第 37 卷 第 13 期
Vol.37 No.13
计 算 机 工 程
Computer Engineering
2011 年 7 月
July 2011

• 开发研究与设计技术 • 文章编号：1000—3428(2011)13—0257—02 文献标识码：A 中图分类号：N945

CUDA 分布式计算的取证应用

刘佳佳，何佳龙，吴顺祥

(厦门大学信息科学与技术学院，福建 厦门 361005)

摘 要：互联网时代犯罪分子多以即时通信软件为工具进行网络诈骗等违法活动。为此，以阿里旺旺为例，研究该工具的数据文件存储结构及其 RC4 和 MD5 加密算法，提出一种基于 CUDA 的分布式计算模型。对 RC4 和 MD5 加密算法采用穷举所有可能密钥的方法进行暴力破解。实验结果表明，该计算模型能准确快速地完成阿里旺旺加密文件的破解，从而为计算机调查取证工作提供技术支持。

关键词：电子商务；CUDA 平台；分布式计算；取证；即时通信软件

Forensics Application of CUDA Distributed Computing

LIU Jia-jia, HE Jia-long, WU Shun-xiang

(School of Information Science and Technology, Xiamen University, Xiamen 361005, China)

【Abstract】In Internet era, a lot of criminals usually carry out Internet fraud and other criminal activities, using instant messaging software as a medium. This paper takes Aliwangwang as an example, and researches into Aliwangwang's information storage structures as well as its data encryption algorithm of RC4 and MD5, proposing a distributed computing model based on CUDA to brute force the data encryption algorithm of RC4 and MD5 by exhausting all possible secret keys. The results show this model can get the right encryption key of Aliwangwang's encrypted file quickly. This research is very useful for computer investigation and forensics in practical use.

【Key words】E-commerce; CUDA platform; distributed computing; forensics; instant messaging software

DOI: 10.3969/j.issn.1000-3428.2011.13.085

进入 21 世纪，即时通信软件使得电子商务活动变得更加方便快捷，但是以此为平台的犯罪活动也随之频繁发生，极大地危害着正常的交易秩序和使用者的正当权益。为此，本文研究一种新的计算机取证方法。

1 相关知识简介

1.1 阿里旺旺

阿里旺旺是淘宝网的主要交易工具，拥有众多客户。其数据文件包含聊天记录和好友列表记录等信息。

1.2 聊天记录文件与好友列表文件

阿里旺旺数据文件的默认存放位置为：{User L}\AliWang Wang\profiles\{User D}\db，其中，User L 表示软件的安装目录；User D 表示阿里旺旺用户名；db 表示后缀是.db 的数据文件；msghis.db 和 user.db 这 2 个文件最重要，msghis.db 是聊天记录文件，user.db 是好友列表文件。

1.3 SQLite 数据库

SQLite 是一个非常小巧的跨平台嵌入式关系数据库，并且源码开放[1]。

SQLite 数据库本身不提供加密功能，但源码中可以找到 2 个预留的加密接口：SQLite3_key 和 SQLite3_rekey，通过这 2 个接口来达到加密的目的[2]。

本文要破解的 msghis.db 和 user.db 文件，就是经 RC4 算法加密后的 SQLite 数据库文件。

1.4 RC4 加密算法

RC4 加密算法算法是一种被广泛使用的高速输出反馈序列加密算法，是可变密钥长度，面向字节操作的流密码。RC4 加密算法主要分密钥排程法、伪随机密钥产生法、产生密文 3 个部分。

2 RC4 加密算法解密

2.1 RC4 算法常用破解方法

文献[3]研究 RC4 算法的引入错误攻击，表明需要 2^{26} 个密钥字和 2^{16} 次错误引入方可恢复 RC4 的初始状态。文献[4]研究了由 FPGA 破解 RC4 算法的理论，指出一个含有 500 个密钥搜索单元的 FPGA 破解 40 位的 RC4 密码算法需要 40 h。近年来，随着计算机硬件性能的提升，暴力破解方法越来越受到研究者的青睐，其算法思想简单，而且便于实现。

2.2 暴力破解

暴力破解也称穷举攻击，通过穷举所有可能的密钥来进行破解，依据简单的轮环形式工作[5]。

暴力破解 RC4 加密算法，必须知道密文的前几个字节对应的明文，使之成为已知明文的攻击。知道明文后，解密密文得到的结果与已知明文进行比较：若解密结果与明文匹配，那么正确的密钥就已经找到。找到了加密密钥，剩余的信息就可解密[6]。而 SQLite 数据库文件有段相同的头信息即“SQLite format 3”。

暴力破解算法的设计思想如下：

(1)得到需要破解的密文 C：对已知的明文 M 用 K 进行加密得到密文 C。M 是 SQLite 数据库头信息的二进制形式，

基金项目：国家自然科学基金资助项目(60704042)；国家“十一五”科技支撑计划基金资助项目(2007BAK34B04)；航空科学基金资助项目(20080768004)；厦门大学 211 信息创新平台基金资助项目(2009-2011)

作者简介：刘佳佳(1987－)，女，硕士，主研方向：模式识别，智能系统；何佳龙，硕士；吴顺祥，教授

收稿日期：2011-02-10 **E-mail：**djxmu@foxmail.com

可以选择其中的 16 个字节：53 51 4c 69 74 65 20 66 6F 72 6D 61 74 20 33 00。

(2)破解密文 C：用密钥空间的密钥逐个对密文 C 解密，把解密结果与已知明文进行比较。

2.3 基于分布式计算的 RC4 算法暴力破解

本文采用基于 CUDA 的分布式计算[7]对 RC4 加密算法进行暴力破解。硬件选用 NVIDIA Tesla S1070 计算系统。对 3 组文件进行了破解，如表 1 所示。发现 3 组 RC4 加密算法的密钥都是 128 bit。考虑到经过 MD5 算法产生的 MD5 值正好也是 128 bit。由此，需用相同方法找到 MD5 值对应的明文信息。

表 1 RC4 算法暴力破解所求密钥

文件序列	密钥	所用时间/min
1	ea96c2a7b194483f32d68bbe0d49b79d	193
2	cf1878c71ff399763cdebf78877b35cb	245
3	0e4eb13887acf8291fb69bd92351bb88	213

2.4 MD5 算法及其解密

对 MD5 算法的简要叙述为：MD5 以 512 位分组来处理输入的信息，且每一分组又被划分为 16 个 32 位子分组，经过了一系列的处理后，算法的输出由 4 个 32 位分组组成，将这 4 个 32 位分组级联后生成一个 128 位散列值[8]。

采用基于 CUDA 的分布式计算技术对表 1 中所求密钥再一次进行暴力破解，算法思想同上。破解结果如表 2 所示。

表 2 MD5 算法暴力破解结果

序列	原始字符串	破解后明文	所用时间/min
1	ea96c2a7b194483f32d68bbe0d49b79d	cntaobaodushijia2008	630
2	cf1878c71ff399763cdebf78877b35cb	cntaobaohaiyuntest	805
3	0e4eb13887acf8291fb69bd92351bb88	cntaobaowangzhi007	684

可以看出，这些 128 位的字符串经暴力破解后显示的明文信息由阿里旺旺的用户名组成。由此，得知阿里旺旺的用户数据文件所采用的加密方式：首先获取到阿里旺旺的用户名，然后求得用户名相应的 MD5 值，以此 MD5 值作为 RC4 算法的密钥对原数据文件进行加密。

2.5 取证系统实用算法

取证系统工作的流程如图 1 所示。

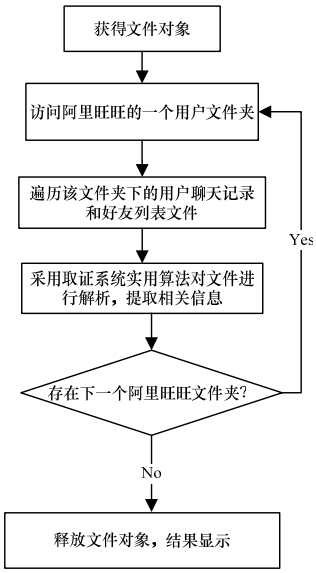

图 1 取证系统流程

取证系统的实用算法描述如下：

Step1 根据聊天记录文件和好友列表文件所在目录获取到阿里旺旺的用户名。

Step2 用 MD5 加密算法计算此用户名的 MD5 值，得到 RC4 加密算法的密钥。

Step3 应用上述密钥，采用 Windows API 函数对加密文件进行解密，得到解密后的聊天记录文件和好友列表文件。

Step4 解密后的文件是 SQLite 数据库文件，通过数据库访问接口对其进行数据读取，以获取聊天记录和好友列表信息。

3 系统测试与验证

本文在 Microsoft Visual Studio 2005 开发环境下，采用 C++语言实现了上述取证系统，并进行了系统性能测试。测试结果显示，此取证系统能够快速、准确地获取到阿里旺旺的聊天记录和好友列表信息。系统运行界面如图 2 所示。

图 2 取证系统运行结果

4 结束语

本文在深入分析即时通信软件下所保存的数据文件的基础上，研究聊天记录和好友列表记录的加密方式，提出基于 CUDA 的分布式计算模型，并成功对淘宝网即时通信软件阿里旺旺进行了破解。实现一个调查取证系统，并优化了输出窗口的树状结构显示效果，使该系统简明全面易操作，为电子商务犯罪的计算机调查取证提供有利帮助。

参考文献

[1] 赵跃华, 朱伟玲. 基于 SQLite 数据库加密模块的设计与实现[J]. 计算机工程与设计, 2008, 29(16): 4132-4134.

[2] 廖顺和, 乐嘉锦. 嵌入式数据库 SQLite 加密方法分析与研究[J]. 计算机应用与软件, 2008, 25(10): 70-71.

[3] 杜育松, 沈 静. 对 RC4 算法的错误引入攻击研究[J]. 电子科学技术大学学报, 2009, 38(2): 253-257.

[4] Nathaniel C, Kenneth B. The Effectiveness of Brute Force Attacks on RC4[C]//Proc. of the 2nd Annual Conf. on Communication Networks and Services Research. [S. l.]: IEEE Press, 2004.

[5] 张丽丽, 张玉清. 基于分布式计算的 RC4 加密算法的暴力破解[J]. 计算机工程与科学, 2008, 30(7): 15-18.

[6] 张丽丽, 张玉清. 基于分布式计算的暴力破解分组密码算法[J]. 计算机工程, 2008, 34(13): 121-123.

[7] 肖 江, 胡柯良, 邓元勇. 基于 CUDA 的矩阵乘法和 FFT 性能测试[J]. 计算机工程, 2009, 35(10): 7-10.

[8] 张裔智, 赵 毅, 汤小斌. MD5 算法研究[J]. 计算机科学, 2008, 35(7): 295-297.

编辑 陈 文